**«Цены на туры»**

**Описание задачи**:

Вам необходимо разработать прогноз продаж для туроператора, продающего туры в Объединенные Арабские Эмираты с помощью статистических методов.

**Дано**:

1. Туроператор «ОАЭ Тур» продает туры в ОАЭ на своем сайте;
2. Единственным каналом продвижения является SEO в Яндексе. Единственным поисковым запросом, по которому оптимизирован сайт, является «туры в оаэ»;
3. У «ОАЭ Тур» есть только один конкурент;
4. Основной фактор конкуренции – это цена;
5. Имеется статистика по средним ценам ну тур у «ОАЭ Тур» и конкурента, а также количеству заказов «ОАЭ Тур», в недельном разрезе.

**Найти**:

1. Определить ключевой рыночный фактор, позволяющий релевантно оценить рыночный спрос на туры в ОАЭ в целом по рынку;
2. Найти значение этого фактора понедельно на весь период наблюдений;
3. Сформировать модель, описывающую зависимость между спросом, ценами и количеством заказов. Привести обоснование модели;
4. С помощью нее рассчитать понедельный прогноз продаж на тестовый период.

**Решение**:

Предположим, что ключевым рыночным фактором, позволяющим релевантно оценить рыночный спрос на туры в ОАЭ в целом по рынку, является популярность туров в ОАЭ в разное время года, сезонный характер спроса на туры в ОАЭ. Для того, чтобы оценить рыночной спрос на туры в ОАЭ, воспользуемся статистикой Яндекса по запросу «туры в ОАЭ» в течение года (информация взята из сервиса Яндекс подбор слов <https://wordstat.yandex.ru/> - История показов по фразе «туры +в оаэ», понедельно):

|  |  |
| --- | --- |
| **Период продаж** | **История показов** |
| 12.03.2018 - 18.03.2018 | 6382 |
| 19.03.2018 - 25.03.2018 | 5218 |
| 26.03.2018 - 01.04.2018 | 4986 |
| 02.04.2018 - 08.04.2018 | 4523 |
| 09.04.2018 - 15.04.2018 | 4105 |
| 16.04.2018 - 22.04.2018 | 3327 |
| **Период продаж** | **История показов** |
| 23.04.2018 - 29.04.2018 | 2546 |
| 30.04.2018 - 06.05.2018 | 2703 |
| 07.05.2018 - 13.05.2018 | 2490 |
| 14.05.2018 - 20.05.2018 | 2644 |
| 21.05.2018 - 27.05.2018 | 2533 |
| 28.05.2018 - 03.06.2018 | 2842 |
| 04.06.2018 - 10.06.2018 | 2822 |
| 11.06.2018 - 17.06.2018 | 2772 |
| 18.06.2018 - 24.06.2018 | 2880 |
| 25.06.2018 - 01.07.2018 | 3007 |
| 02.07.2018 - 08.07.2018 | 3697 |
| 09.07.2018 - 15.07.2018 | 4001 |
| 16.07.2018 - 22.07.2018 | 4083 |
| 23.07.2018 - 29.07.2018 | 4406 |
| 30.07.2018 - 05.08.2018 | 4845 |
| 06.08.2018 - 12.08.2018 | 5482 |
| 13.08.2018 - 19.08.2018 | 5489 |
| 20.08.2018 - 26.08.2018 | 5481 |
| 27.08.2018 - 02.09.2018 | 5244 |
| 03.09.2018 - 09.09.2018 | 5980 |
| 10.09.2018 - 16.09.2018 | 6604 |
| 17.09.2018 - 23.09.2018 | 7975 |
| 24.09.2018 - 30.09.2018 | 8237 |
| 01.10.2018 - 07.10.2018 | 9068 |
| 08.10.2018 - 14.10.2018 | 8850 |
| 15.10.2018 - 21.10.2018 | 8339 |
| 22.10.2018 - 28.10.2018 | 8180 |
| 29.10.2018 - 04.11.2018 | 7836 |
| 05.11.2018 - 11.11.2018 | 7509 |
| 12.11.2018 - 18.11.2018 | 7502 |
| 19.11.2018 - 25.11.2018 | 6198 |
| 26.11.2018 - 02.12.2018 | 5903 |
| 03.12.2018 - 09.12.2018 | 6325 |
| 10.12.2018 - 16.12.2018 | 5740 |
| 17.12.2018 - 23.12.2018 | 5315 |
| 24.12.2018 - 30.12.2018 | 4551 |
| 31.12.2018 - 06.01.2019 | 5270 |
| 07.01.2019 - 13.01.2019 | 7902 |
| 14.01.2019 - 20.01.2019 | 8145 |
| 21.01.2019 - 27.01.2019 | 7169 |
| 28.01.2019 - 03.02.2019 | 6908 |
| 04.02.2019 - 10.02.2019 | 7193 |
| 11.02.2019 - 17.02.2019 | 6620 |
| 18.02.2019 - 24.02.2019 | 6520 |
| 25.02.2019 - 03.03.2019 | 6650 |
| 04.03.2019 - 10.03.2019 | 5393 |

Для того, чтобы рассчитать понедельный прогноз продаж на тестовый период, воспользуемся моделью множественной линейной регрессии. Перед тем, как начать расчет коэффициентов, посмотрим на факторы и на целевую переменную, которые войдут в финальную модель:

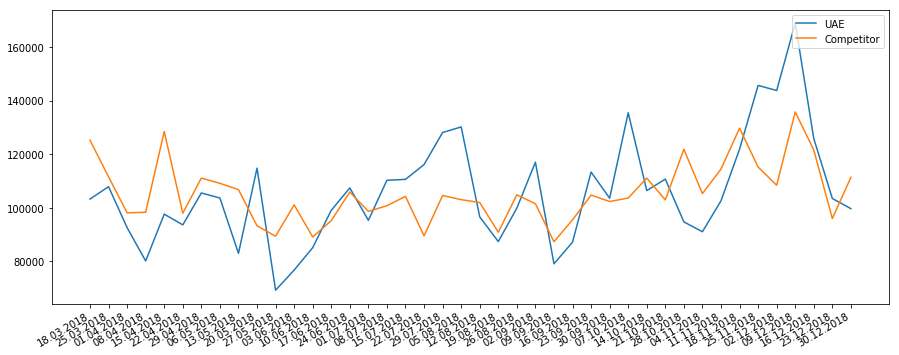
**Факторы**

1. Средняя цена «ОАЭ ТУР», руб./заказ
2. Средняя цена конкурента, руб./заказ
3. История показов, шт.

**Целевая переменная**

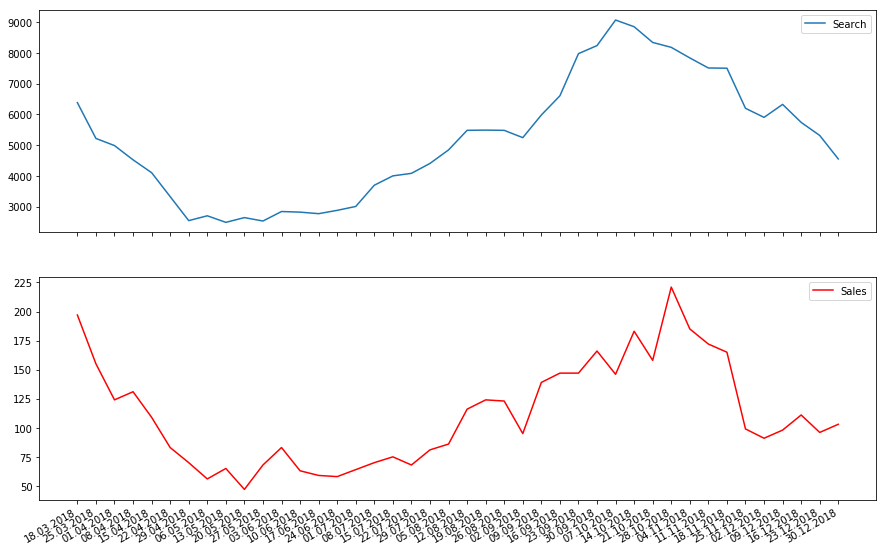
1. Факт продаж, заказов

Посмотрим на графики средних цен туров у оператора «ОАЭ ТУР» и его конкурента:



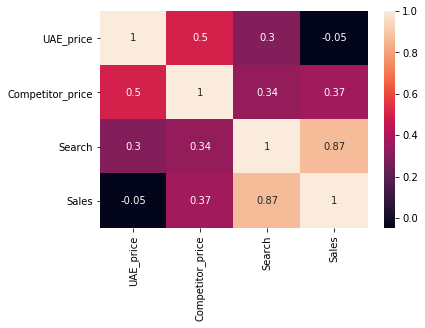
Оба ряда имеют схожую динамику, однако график средних цен «ОАЭ ТУР» более волатилен. Вероятно, два этих фактора коррелированы друг с другом.

Посмотрим на графики факторов «История показов» и «факт продаж»:



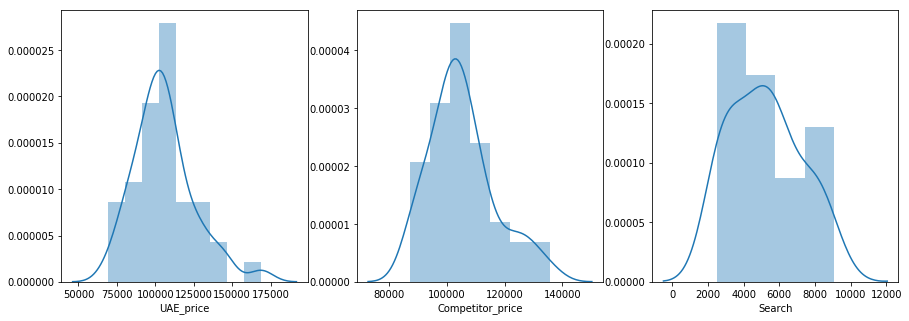
Информация на графиках позволяет сделать вывод о сезонном характере спроса на туры в ОАЭ: спрос растет в октябре-ноябре и падает в мае-июне. Интерес к турам в ОАЭ также наблюдается в начале года. Фактор «История запросов», вероятно, дает достаточно хорошую оценку спросов на туры в ОАЭ в рассматриваемом периоде.

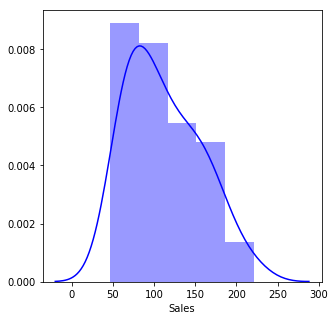
Проведем корреляционный анализ:



Как и было предположено ранее, факторы «Средняя цена «ОАЭ ТУР» и «Средняя цена конкурента» имеют достаточно большой коэффициент корреляции *r = 0,5.* Фактором, имеющим наибольшую корреляцию с целевой переменной, является фактор «История показов» - *r = 0,87.*

Посмотрим на распределения:





Переменные «Средняя цена «ОАЭ ТУР» и «Средняя цена конкурента» имеют распределение, схожее с нормальным. Распределения переменных «История показов» и «Факт продаж» имеют положительную асимметрию и вытянуты вправо. В данных отсутствуют большие выбросы.

Построим модель множественной линейной регрессии. Для анализа качества модели разделим выборку на тренировочную и тестовую. Коэффициенты модели:

|  |  |  |  |
| --- | --- | --- | --- |
| constant | b1 | b2 | b3 |
| -24.194 | -0.00119 | 0.00146 | 0.0210 |

Качество на тренировочной выборке:

The model performance for training set

--------------------------------------

RMSE is 11.771799116309134

R2 score is 0.9304481378061722

Качество на тестовой выборке:

The model performance for testing set

--------------------------------------

RMSE is 12.728027420703693

R2 score is 0.9054590434218177

Качество в кросс-валидации:

Train R2 [0.93707853 0.71734108 0.94318277 0.93412645 0.89745198]

Test R2 [0.92986665 0.94903511 0.96806118 0.9524406 0.90964487]

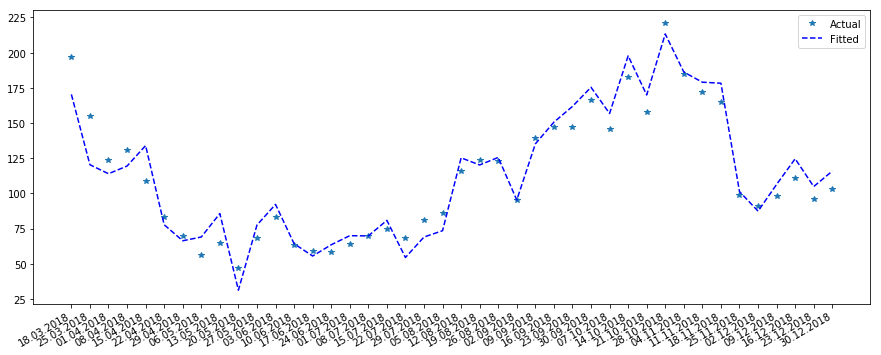
Статистика модели:

|  |  |  |  |
| --- | --- | --- | --- |
| **Dep. Variable:** | Sales | **R-squared:** | 0.930 |
| **Model:** | OLS | **Adj. R-squared:** | 0.922 |
| **Method:** | Least Squares | **F-statistic:** | 107.0 |
| **Date:** | Thu, 02 May 2019 | **Prob (F-statistic):** | 5.00e-14 |
| **Time:** | 21:06:31 | **Log-Likelihood:** | -108.77 |
| **No. Observations:** | 28 | **AIC:** | 225.5 |
| **Df Residuals:** | 24 | **BIC:** | 230.9 |
| **Df Model:** | 3 |  | |
| **Covariance Type:** | nonrobust |

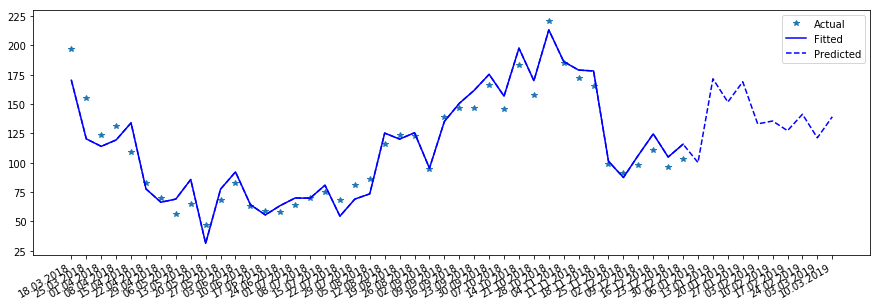
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **t** | **P>|t|** | **[0.025** | **0.975]** |
| **const** | -24.12 | 23.184 | -1.044 | 0.307 | -72.04 | 23.656 |
| **UAE\_price** | -0.0012 | 0.000 | -8.594 | 0.000 | -0.001 | -0.001 |
| **Competitor\_price** | 0.0015 | 0.000 | 5.184 | 0.000 | 0.001 | 0.002 |
| **Search** | 0.0210 | 0.002 | 13.462 | 0.000 | 0.018 | 0.024 |

Значения p-value при коэффициентах позволяют отвергнуть нулевую гипотезу о незначимости отличия оценок регрессионных коэффициентов от нуля. При этом, значение p-value при константе эту гипотезу отвергнуть не позволяет, однако из-за природы константы этим фактом можно пренебречь. Значение F-критерия также говорит о значимости всех регрессоров в целом.

Модель показала хорошие результаты: на тренировочной и тестовой выборках коэффициент детерминации составил более 0.9, что позволяет сделать вывод о том, что модель передает более 90% изменчивости исходного ряда. Модель обладает достаточной устойчивостью: значения R2 на разных фолдах в кросс-валидации незначительно отклоняются от среднего.



С помощью модели рассчитаем понедельный прогноз продаж на тестовый период:



|  |  |  |  |
| --- | --- | --- | --- |
| Период продаж | Средняя цена «ОАЭ ТУР», руб./заказ | Средняя цена конкурента, руб./заказ | Прогноз продаж, заказов |
| 31.12.2018 - 06.01.2019 | 106651,7 | 96013,79 | 95 |
| 07.01.2019 - 13.01.2019 | 101673,5 | 102984,4 | 153 |
| 14.01.2019 - 20.01.2019 | 123099,7 | 103358,1 | 132 |
| 21.01.2019 - 27.01.2019 | 119102,7 | 125948,5 | 146 |
| 28.01.2019 - 03.02.2019 | 101896,1 | 91143,04 | 122 |
| 04.02.2019 - 10.02.2019 | 124413,5 | 107045,3 | 119 |
| 11.02.2019 - 17.02.2019 | 125538,7 | 110561,7 | 112 |
| 18.02.2019 - 24.02.2019 | 111693,3 | 110301,8 | 126 |
| 25.02.2019 - 03.03.2019 | 112086,9 | 95005,36 | 117 |
| 04.03.2019 - 10.03.2019 | 85283,17 | 103556 | 127 |